1 Supplementary

In this supplementary material, we cover a generalized description of the Gaussian Process and the algorithm of the N4SID technique.

1.1 Gaussian Process

A Gaussian process generalizes multivariate normal distributions to distributions over functions that are specified by a prior mean function $\mu(d)$ and covariance function $\kappa(d_i, d_j; \theta)$ as

$$f(d) \sim \mathcal{GP}(\mu(d), \kappa(d_i, d_j; \theta)) \tag{1}$$

Where θ indicates the kernel hyper-parameters that need to be optimized. The predictive distribution of the GP evaluated at the test points d_* of data $\mathcal{D} = \{(d_i, y_i)\}_{i=1}^n$, is given by

$$p(f(d_*)|\mathcal{D}) \sim \mathcal{GP}(\mu_{f|\mathcal{D}}(d_*), \kappa_{f|\mathcal{D}}(d_*, d'_*)), \tag{2a}$$

$$\mu_{f|\mathcal{D}}(d) = \mu(d_*) + K_{d_* d} \hat{K}_\theta^{-1} y, \tag{2b}$$

$$\kappa_{f|\mathcal{D}}(d_*, d'_*)) = K_{d_* d_*} - K_{d_* d} \hat{K}_\theta^{-1} K_{d_* d}^T \tag{2c}$$

where, $\hat{K}_\theta = K_\theta + \sigma^2 \mathrm{I}$ is the covariance (kernel) matrix for the noisy targets y, and $y = (y(x_1), ..., y(x_n))^T$. All kernel matrices implicitly depend on hyper-parameters, θ, and the log marginal likelihood conditioned on the hyper-parameters is given by

$$\log p(y|\theta) = -\frac{1}{2} y^T \hat{K}_\theta^{-1} y - \frac{1}{2} \log |\hat{K}_\theta| - \frac{n}{2} \log 2\pi \tag{3}$$

where the three terms of the eq.3 have readily interpretable roles: $-\frac{1}{2} y^T \hat{K}_\theta^{-1}$is the data-fit which involves the observed targets; $\frac{1}{2} \log |\hat{K}_\theta|$ is the complexity penalty depending only on the kernel function and the inputs and $\frac{n}{2} \log 2\pi$ is a normalization constant. Maximizing the log marginal likelihood provides a utility function for kernel learning.

1.2 N4SID System Identification Algorithm

In this work, we implement an alternative technique known as N4SID, a data-driven approach designed for system identification (proposed by Van Overschee & De Moor). This method utilizes a subspace-based strategy, which segregates the data into deterministic and stochastic elements by projecting them onto distinct orthogonal subspaces. The algorithm calculates the system's state sequence, state-transition matrid, input matrid, and output matrix from these subspaces, offering a concise and effective depiction of the inherent dynamical system. Essentially, N4SID is capable of directly approximating system behavior using input-output data in a state-space format. A description of the algorithm is shown below:

Algorithm 1 N4SID Algorithm

```
1: procedure N4SID
2:     Normalize input-output data.
3:     Estimate a covariance matrix by applying QR decomposition.
4:     Compute a singular value decomposition (SVD) of the covariance matrix.
5:     Divide the SVD output into observable and unobservable subspaces.
6:     From the observable subspace, compute the system matrices A, B, C, and D.
7:     Perform the balancing transformation and reduction of the state-space model.
8:     Return the state-space model.
9: end procedure
```

In this paper, considering the collected input-output data set $u(k), y(k)_{k=1}^N$, with $u(k) \in \mathbb{R}^m$ representing the input data sample,$\{v_{xk}, v_{yk}, \psi_k, \delta_k, \omega_k, a_{xk}, \Delta\delta_k, u_{Tk}, u_{Bk}\}$, and $y(k) \in \mathbb{R}^p$ denoting the output vector, $\epsilon_{v_{x\,k}}, \epsilon_{v_{y\,k}}, \epsilon_{\omega_k}$, at time k, N4SID can effectively approximate the system's dynamics using a state-space representation, as demonstrated in 4.

$$\begin{cases} x(t+1) = Ax(k) + Bu(k) \\ y(t) = Cx(k) + Du(k) \end{cases} \tag{4}$$